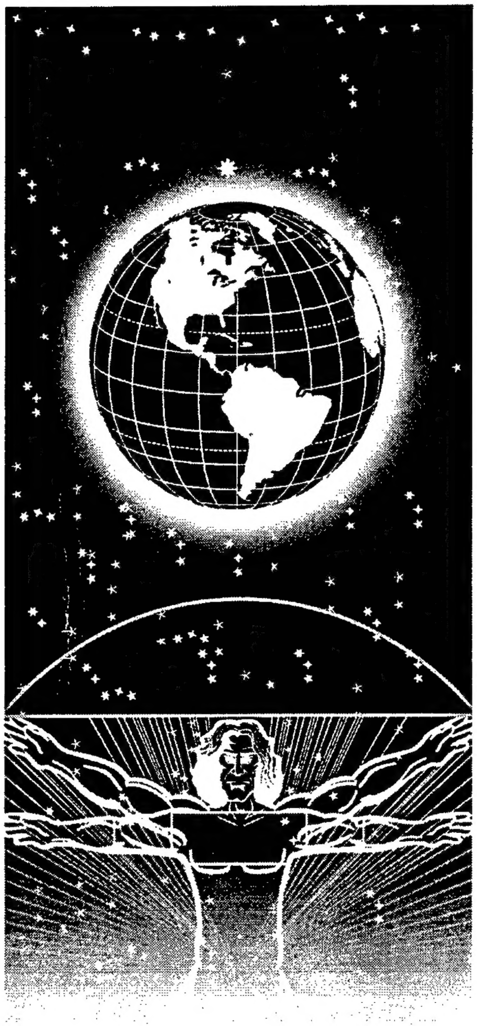
AFRL-HE-BR-TR-1999-0033



UNITED STATES AIR FORCE RESEARCH LABORATORY

MODERATORS OF SKILL RETENTION INTERVAL-PERFORMANCE DECREMENT RELATIONSHIPS: IMPLICATIONS FOR SKILL MAINTENANCE

Charles E. Lance
Anthony G. Parisi
Department of Psychology
The University of Georgia
Athens, GA 30602-3013

Winston Bennett, Jr.
HUMAN EFFECTIVENESS DIRECTORATE
MISSION CRITICAL SKILLS DIVISION
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

Mark Teachout
United States Automobile Association

Donald L. Harville
HUMAN EFFECTIVENESS DIRECTORATE
WARFIGHTER TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

Marsha L. Welles
Department of Psychology
The University of Georgia
Athens, GA 30602-3013

May 1998

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
HUMAN EFFECTIVENESS DIRECTORATE
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

NOTICES

This report is published in the interest of scientific and technical information exchange and does not constitute approval or disapproval of its ideas or findings.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

WINSTON BENNETT, JR
Project Scientist

R. BRUCE GOULD
Acting Chief
Mission Critical Skills Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
May 1998

3. REPORT TYPE AND DATES COVERED
Interim Report-September 1996-December 1997

4. TITLE AND SUBTITLE
Moderators of Skill Retention Interval-Performance Decrement Relationships: Implications for Skill Maintenance

5. FUNDING NUMBERS
C- F41624-97-C-5030
PE- 62202F
PR- 1123
TA- A2
WU 20

6. AUTHOR(S)
Charles E. Lance
Anthony G. Parisi
Winston Bennett, Jr.
Mark Teachout
Donald L. Harville
Marsha I. Welles

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Psychology
The University of Georgia
Athens, GA 30602-3013

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Mission Critical Skills Division
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-BR-TR-1999-0033

11. SUPPLEMENTARY NOTES
Air Force Research Laboratory Technical Monitor: Winston Bennett (480) 988-6561, DSN 474-6297

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A robust finding in the literature on skill maintenance is that extended skill retention intervals lead to decreased task performance. We investigated three potential moderators of the skill retention interval (SRI) - task performance decrement relationship: (a) degree of initital skill learning, (b) performer aptitude, and (c) degree of task difficulty. Results in eight samples of enlisted U.S. Air Force personnel (total N = 1,544) strongly supported a negative SRI - task performance relationship, but support for moderator effects was mixed. Implications for skill maintenance in actual work settings are discussed.

14. SUBJECT TERMS
Skill Retention Interval
Moderators
Job Experience
Task Difficulty
Aptitude

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z-39-18
298-102 *COMPUTER GENERATED*

Table of Contents

LIST OF FIGURES

LIST OF TABLES

PREFACE

The work described in this paper was conducted under contract F41622-95-P-4979, with the Air Force Armstrong Laboratory Human Resources Directorate, Technical Training Research Division.

The views and opinions expressed in this paper are those of the authors and do not reflect the official policies or opinions of their respective organizations.

An earlier version of this paper was presented at the April, 1996 meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

The authors wish to thank the United States Air Force personnel who participated in this project, for these scientific advances would not be possible without their support and cooperation. The authors would also like to thank Kathleen Sheehan for preparing the final format of this report.

SUMMARY

A robust finding in the literature on skill maintenance is that extended skill retention intervals lead to decreased task performance. We investigated three potential moderators of the skill retention interval (SRI) - task performance decrement relationship: a) degree of initial skill learning, b) performer aptitude, and c) degree of task difficulty. Results in eight samples of enlisted U.S. Air Force personnel (total $\underline{N} = 1{,}544$) strongly supported a negative SRI - task performance relationship, but support for moderator effects was mixed. Implications for skill maintenance in actual work settings are discussed.

MODERATORS OF SKILL RETENTION INTERVAL - PERFORMANCE DECREMENT RELATIONSHIPS: IMPLICATIONS FOR SKILL MAINTENANCE

INTRODUCTION

It is not unusual for employees to receive training on skills that they may not be called upon to use for some time later. Police officers may go weeks or months without firing a weapon or apprehending a dangerous felon. Disaster teams may go years without evacuating residents from affected areas, managing evacuation routes, and rescuing survivors. Major wars are, thankfully, infrequent. Yet when military reserve personnel are called into active duty they are expected to perform their missions effectively (Wisher, Sabol, Sukenik, Hillel, & Kern, 1991). Each of these scenarios depicts a situation in which significant performance decrements may incur over extended periods of skill non-use. The purpose of this paper is to investigate, in realistic field settings, factors which may mitigate against performance decrements over extended skill retention intervals.

Background

Compared to the abundant literatures that exist on (a) learning and learning strategies (e.g., Ackerman, 1987, 1992; Ackerman & Humphreys, 1991; Gallagher, 1994; Hintzman, 1990; Weiss, 1991), (b) training program design, training techniques, and training evaluation methodologies (Goldstein, 1989, 1991; Kraiger, Ford, & Salas, 1993; Snow & Swanson, 1992; Tannenbaum & Yukl, 1992), and (c) transfer of training (Baldwin & Ford, 1988; Gagne, Baker, & Foster, 1950; Michalak, 1981; Royer, 1979; Schmidt & Bjork, 1992), little has been written on the maintenance of job skills over extended skill retention intervals. Nonetheless, there are reviews of the empirical skill maintenance literature that does exist, including Arthur, Bennett, McNelly, and Stanush (1995), Farr (1987), Fendrich, Healy, Meiskey, Crutcher, Little, and Bourne (1988), Hagman and Rose (1983), Naylor and Briggs (1961), and Schendel, Shields, and Katz (1978).

One of the more robust empirical findings in this literature, which consists primarily of laboratory studies, is of a positive relationship between the length of the skill retention interval (i.e., the period of skill non-use) and decreased task performance (Arthur et al., 1995; Farr, 1987; Naylor & Briggs, 1961; Prophet, 1977). That is, evidence indicates that the length of the skill retention interval (SRI) is negatively related to task proficiency (Arthur et al., 1995; Fendrich et al., 1988; Mengelkoch, Adams, & Gainer, 1971; Schendel & Hagman, 1982). The focus of this paper is on factors which may mitigate against performance decrements over extended SRIs, that is, potential moderators of the SRI - performance decrement relationship.

SRI - Performance Decrement Moderators

Some research has been directed toward identifying factors that mitigate against performance decrements over extended SRIs. For example, Schendel and Hagman (1982) investigated two possible methods: (a) increasing initial task overlearning, and (b) providing refresher training, on performance of a task involving the disassembly/assembly of an M60 machine gun. Results indicated that the overtraining group performed better than the refresher training group and a control group. A subsequent meta-analysis by Driskell, Willis, and Copper (1992) confirmed the beneficial effects of overlearning in mitigating against performance decrements over extended SRIs. These findings, along with Farr's (1987) review, indicate that the degree of initial skill learning appears to moderate the SRI - task performance relationship: performance on well-learned tasks tends to decline less rapidly over extended SRIs than does performance on less well learned tasks.

Other reviews (e.g., Arthur et al., 1995; Farr, 1987) suggest additional potential moderators of the negative relationship between length of the SRI and task performance, including:

Task complexity. Although a number of task characteristics have been suggested as possibly being related to skill maintenance (e.g., continuous-control vs. procedural tasks; motor vs. verbal tasks; naturally occurring vs. contrived tasks; Arthur et al., 1995), task complexity or difficulty has been hypothesized to be one of the more potent factors affecting skill maintenance and subsequent task performance (Farr, 1987). Theoretically, those skills that are the most difficult to learn may also be those that are most easily forgotten (Mumford, Weeks, Harding, & Fleishman, 1987). Thus, the hypothesis here was of greater task performance decrements for more difficult tasks as compared to less difficult ones over extended SRIs.

Ability. Individual cognitive ability is positively related to skill acquisition (Herrnstein & Murray, 1994). This is demonstrated in actual work settings by the consistent relationship between general cognitive ability and success in job training (e.g., Carretta & Ree, 1995, Earles & Ree, 1992; Olea & Ree, 1994; Ree, Carretta, & Teachout, 1995; Ree & Earles, 1991a). Some research has demonstrated a compensatory (additive) relationship between cognitive ability and job experience (e.g., Schmidt, Hunter, Outerbridge, & Goff, 1988). Thus there may be multiple paths to a given level of skill proficiency as a function of ability versus experience tradeoffs. However, all else equal (including present skill levels), lower ability performers will likely forget learned skills more quickly than will higher ability performers (Farr, 1987). Thus it was hypothesized that task performance will be less adversely affected over extended SRIs for higher- as compared to lower-ability performers.

Purpose

One of the major limitations of the existing research on skill maintenance is that the majority of it has been conducted in the laboratory, over relatively short skill retention intervals, and on contrived experimental tasks (e.g., Fendrich et al., 1988). We do not fault previous skill maintenance research on these grounds, as (a) laboratory studies usually permit stronger causal inferences than do field studies (Dobbins, Lane, & Steiner, 1988; James, Mulaik, & Brett, 1982), and (b) many of the most important developments in applied psychology are rooted in basic research (e.g., Adams, 1972). However, basic research must be complemented with applied field research to corroborate and establish the external validity of basic research findings obtained in the laboratory (Cook & Campbell, 1979; Cook, Campbell, & Peracchio, 1990; Driskell & Salas, 1992; Latham & Lee, 1986). Thus in the present context, additional field research is needed to determine the extent to which (mainly laboratory) findings summarized by the major reviews on skill maintenance and task performance generalize to actual work settings. This points to the purpose of the present study, which was to test hypothesized moderating effects of (a) degree of initial skill learning, (b) task complexity/difficulty, and (c) learner cognitive ability, on the relationship between the length of the SRI and decrements in task performance levels, in actual work-related settings.

METHOD

Samples

Data reported here were collected by the Air Force Human Resources Laboratory (AFHRL) in the late 1980s as part of the Joint-Service Job Performance Measurement (JPM)/Enlistment Standards Project (Wigdor & Green, 1986, 1991). A number of criteria were used to select representative Air Force Specialties (AFSs) for the U.S. Air Force (USAF) JPM project, including the number of first-term airmen populating various AFSs in the then-current USAF enlisted occupational structure, level of AFS occupational learning difficulty, and Armed Services Vocational Aptitude Battery (ASVAB) composite area assignment (i.e., Mechanical, Administrative, General, or Electronic (MAGE); Department of Defense, 1984). Data reported here were collected from airmen in the eight AFSs chosen for the USAF JPM project as being representative of relatively populous AFSs, broad aptitude requirement areas, and a range of occupational difficulties. These are listed in Table 1.

Procedure

The conceptualization, design, and implementation of the USAF JPM project is described in detail in a number of previous publications (e.g., Hedge & Teachout, 1986, 1992; Kavanagh, Borman, Hedge, & Gould, 1987; Lance, Teachout, & Donnelly, 1992; Laue, Hedge, Wall, Pederson, & Bentley, 1992; Ree, Earles, & Teachout, 1994; Teachout & Pellum, 1991), and so

are not discussed in great detail here. Briefly, each airman selected for participation in the JPM project was assessed in a Walk-Through Performance Test (WTPT) developed specifically for their AFS. The WTPT consisted of a number of task items which the participant either performed in a hands-on mode (i.e., as in a work sample) or in a "show-and-tell" interview mode. For each AFS, and in order to reduce the possibility of ceiling effects in WTPT performance, tasks were sampled so that 40% of the tasks included in the WTPT were from the 4th (i.e., most difficult) quartile of learning difficulty, 30% were from the 3rd quartile, 20% from the 2nd quartile, and 10% were from the 1st (easiest) quartile. Performance on each WTPT task item was scored as a weighted sum of the number of task steps completed correctly as recorded by the test administrator. Additionally, participants indicated (a) the number of times they had performed each task on the job prior to its being administered in the WTPT ("Number of Times Performed"), and (b) how long (in weeks) it had been since they had last performed the task ("Last Time Performed"). Finally, each participant rated their prior relative experience in performing each task on the job (1 = No or almost no experience, to 7 = A very great amount of experience).

TABLE 1: Air Force Specialties Selected for USAF JPM Data Collection

AFSC	N	MAGE Area	Title
122x0	231	General	Aircrew Life Support Specialist
272x0	191	General	Air Traffic Control Operator
324x0	140	Electronics	Precision Measuring Equipment Specialist
328x0	98	Electronics	Avionic Communications Specialist
423x5	272	Mechanical	Aerospace Ground Equipment Specialist
426x2	255	Mechanical	Jet Engine Mechanic
492x1	156	Administrative	Information Systems Radio Operator
732x0	201	Administrative	Personnel Specialist

Note. AFSC = Air Force Specialty Code; N = sample size; MAGE Area = MAGE aptitude area assignment (see text); Title = AFS title.

Measures

Task Performance. Task performance was assessed by the WTPT score which was a weighted percentage of task steps completed correctly x 10. Some task items in the WTPT were administered in both the hands-on and interview modes, as mentioned earlier. In these cases, the hands-on score was chosen for analysis. For other tasks, the score was used for whichever mode in which the task was administered. As reported by Laue, Teachout, and Harville (1990), correlations between interview and hands-on items range between .56 and .88 with a median r = .76, supporting their equivalence.

Skill Retention Interval (SRI). SRI was measured using participants' Last Time Performed reports of the number of weeks it had been since they had last performed each task in the WTPT. Thus, SRI reflects the retention interval between the time that participants were tested on each task and the time that they had last practiced it.

Initial Skills Learning. We used measures of job and task experience as surrogate measures of amount of initial skills learning as more direct measures (e.g., technical school training grades) were not available at the task-level of analysis, and where available, would not necessarily reflect the degree of initial skills learning on the particular tasks included in the WTPTs. We used experience indicators of amount of initial skills learning under the rationales that (a) research has shown that additional accrued experience leads to more effective job and task performance (Quinones, Ford, & Teachout, 1995), via enhanced skills developed through experience, and (b) that skill level as manifested in task performance ought to deteriorate less quickly over extended skill retention intervals for those who had accrued greater (as opposed to lesser) previous initial experience levels (and implicitly, skill levels) prior to the onset of the SRI. Thus we assumed that additional job and task skills are acquired with additional job and task experience, and that measures of experience levels prior to the onset of the SRI were suitable surrogate measures for the degree of skill learning at the time that the SRI commenced.

We measured Task Experience as a composite of participants' reports of the number of times they had previously performed each WTPT task ("Number of Times Performed" (NTP)), and their relative task experience ratings (TERs). In each AFS, NTP was markedly skewed and multimodal. To more approximately normalize NTP and to more nearly equate its scale with TERs, we created a transformed NTP (TNTP) as (1 = 0 to 1 NTP, 2 = 2 to 10 NTP, 3 = 11 to 20 NTP, 4 = 21 to 50 NTP, 5 = 51 to 100 NTP, 6 = 101 to 800 NTP, and 7 = 801 to 999 NTP). Task Experience was then computed as the mean of TER and TNTP for each task. Thus the Task Experience composite represents a micro, task-level surrogate measure of degree of initial skill learning.

We computed Job Experience as months of Total Active Federal Military Service. For most participants, this corresponded to the difference (in months) between the WTPT administration date and the beginning date of their current enlistment (i.e., almost all participants were in their first term of enlistment). Thus Job Experience represents a much more macro, job-level surrogate measure of degree of initial skill learning.

Task Difficulty. As part of the USAF's periodic task inventory (Occupational Survey) job analyses of AFSs, data are collected on the learning difficulty of tasks performed by incumbents (see Burtch, Lipscomb, & Wissman, 1982; Mumford et al., 1987; Weeks, 1984). For the present study, we used task learning difficulty indices (LDIs) obtained from the occupational survey conducted just prior to each AFSs' JPM data collection to index task difficulty. Task LDIs range from a low of 1 (very easy tasks) to a high of 9 (very difficult tasks).

Aptitude. We accessed personnel files to obtain each JPM participant's pre-enlistment ASVAB scores. The ASVAB consists of 10 specific ability subtests, each normed to a mean of 50 and standard deviation of 10. Although the ASVAB is a differential aptitude test battery, the Air Force Qualifying Test (AFQT) score [computed from ASVAB subtests as Arithmetic Reasoning + Mathematical Knowledge + 2*(Word Knowledge + Paragraph Comprehension), see Department of Defense, 1984], is regarded as a strong indicator of general cognitive ability (g) (Ree & Carretta, 1994; Ree & Earles, 1991b, Ree et al., 1994). Thus, we used WTPT participants' AFQT percentile scores to reflect g.

Input Data Sets

The input data set for each AFS was "multilevel" in the sense that task performance (WTPT scores) and the predictor measures were operationalized at varying levels of analysis (Lance, Hedge, & Alley, 1989). For example, each task performance (i.e., WTPT) score indexed the ith participant's score on the jth task included in the AFS's WTPT. Thus the number of observations in each data set corresponded to N participants times J tasks (i.e., N*J). SRI and Task Experience measures also varied across each ith participant x jth task combination to indicate (a) how long it had been since the ith participant had performed the jth task, and (b) the amount of previous experience the ith participant had in performing the jth task, respectively. On the other hand, Job Experience and Aptitude scores (two person-level variables) varied appropriately across the N participants, but were constant for each participant across all J tasks that the ith participant completed. Complimentarily, Task Difficulty scores (a task-level variable) varied appropriately across the J tasks, but were constant for all of the N participants performing the task.

Analyses

Our main interest was on the potential moderating effects of Task Difficulty, Aptitude, and initial skill learning (as reflected by Job and Task Experience levels) on the SRI - task performance relationship. Generally, we hypothesized ordinal interactions. For example, we expected that task performance would be less adversely affected over extended SRIs for less difficult, as compared to more difficult tasks. Thus, we expected a generally negative relationship between WTPT and SRI scores (indicating the general tendency for performance to decline over extended SRIs), as well as moderator effects of task difficulty, initial skill learning, and performer cognitive ability on the WTPT - SRI relationship. Thus, we first examined the zero-order correlation between WTPT and SRI to confirm the expected performance decline over longer retention intervals.

Next we tested moderator hypotheses using hierarchical moderated regression (Cohen & Cohen, 1975) by first forming cross-products between SRI scores and each of the hypothesized moderators. We then regressed WTPT on SRI and the hypothesized moderator (e.g., SRI and Task Difficulty) in the first-step regression, and included the appropriate cross-product (e.g., SRI x Task Difficulty) in the second step. A statistically significant change in the $\underline{R}^2$s between the first and second-step regressions indicated the presence of a significant moderator effect.

RESULTS

Table 2 shows descriptive statistics for study variables.[1] Mean WTPT scores were well below the maximum of 10.00, indicating the absence of ceiling effects in performance. WTPT SDs ranged between 2.34 and 4.20, also suggesting that restriction of range was not a problem. Mean SRI scores varied considerably across AFSs from a low of 6.20 weeks for the Information Systems Radio Operator sample to a high of almost 30 weeks ($\underline{M}$ = 29.63) for the Aircrew Life Support sample. Also, SRI scores were highly positively skewed (range 1.81 to 5.82). To reduce possible biasing effects of this positive skew, we used the natural log transform (LN) of SRI in regression analyses. As Table 2 shows taking LN(SRI) greatly reduced, but did not entirely eliminate skewness.

Mean Task Difficulty values in Table 2 show that tasks included in the WTPTs were moderately difficult, on the average. This reflects the fact that WTPTs were designed to include a greater proportion of difficult, as compared to easier, tasks (see Laue et al., 1992). Task Experience means indicated that, on the average, participants were moderately familiar with the tasks they performed in the WTPTs. Job Experience means ranged from slightly less than two years (AFSC 492x1) to just under three years (AFSC 328x0). Finally, mean Aptitude scores all were above the 50th percentile and in two AFSs (AFSCs 324x0 and 328x0) they were close to the 80th percentile. This reflects effects of preselection of participants on the basis of ASVAB scores.[2]

TABLE 2: Descriptive Statistics for Study Variables

Variable		122x0	272x0	324x0	328x0	423x5	426x2	492x1	732x0
		Air Force Specialty Code (AFSC)							
WTPT	Mean	7.07	6.77	7.69	7.44	5.32	6.92	6.09	7.69
	S. D.	2.86	2.58	2.91	3.02	2.92	2.34	4.20	3.16
	Skew	-0.98	-0.65	-1.15	-0.95	-0.28	-0.71	-0.39	-1.25
SRI	Mean	29.58	8.06	16.70	11.51	24.32	11.09	6.20	6.71
	S. D.	35.86	14.91	25.20	20.15	33.20	17.47	14.67	16.37
	Skew	1.81	3.31	2.53	3.18	2.84	3.60	5.82	4.94
LN (SRI)	Mean	4.67	1.49	2.05	1.77	2.45	1.83	1.27	1.26
	S. D.	2.97	1.04	1.26	1.12	1.29	1.09	0.98	0.99
	Skew	0.77	1.20	0.48	0.86	0.19	0.59	1.42	1.77
Task Difficulty	Mean	4.81	4.72	4.56	4.86	5.02	4.96	3.89	4.29
	S. D.	1.04	0.63	0.84	1.20	0.90	0.67	0.68	1.09
	Skew	0.27	-0.09	0.28	0.06	0.31	0.60	0.38	-0.36
Task Experience	Mean	3.99	4.58	3.71	4.23	3.36	3.64	4.95	4.84
	S. D.	1.73	1.95	1.68	1.50	1.45	1.64	1.54	1.83
	Skew	-0.07	-0.32	0.15	-0.12	0.33	0.27	-0.49	-0.58
Job Experience	Mean	29.26	26.15	27.44	34.81	27.95	31.11	22.91	28.04
	S. D.	10.93	9.59	10.41	15.24	10.31	11.95	12.79	11.51
	Skew	-0.25	0.08	0.50	0.23	0.23	0.29	0.65	0.33
Aptitude	Mean	59.24	72.49	78.47	78.24	56.93	55.39	60.78	61.51
	S. D.	16.65	16.05	14.99	14.60	15.41	19.49	16.81	16.20
	Skew	0.36	-0.10	-0.64	-0.64	0.51	0.35	0.47	0.42

Note. WTPT = Walk-Through Performance Test; SRI = Skill retention interval. See Table 1 for AFSs' Titles.

Table 3 summarizes correlational and hierarchical regression results for each of the eight AFSs. The first column indicates the number of observations (i.e., N*J), or the effective sample size for each AFS (once missing data were deleted). The second column shows the zero-order rs between task performance (WTPT scores) and natural log-transformed SRI scores which indicate the length of the skill retention interval. The WTPT - LN(SRI) relationship was statistically significant in all eight AFSs. Inexplicably however, the direction of the WTPT - LN(SRI) relationship was opposite to that predicted in one sample (AFSC 492x1). However, in the remaining samples there was a significant negative relationship between task performance (WTPT scores) and extended SRIs, as was predicted.

Tests of the moderator hypotheses are summarized in the remaining columns in Table 3, where the R^2 changes from the first- to second-step hierarchical regressions and associated probabilities are shown. Results in Table 3 show mixed support for moderator hypotheses. Task Experience moderated the WTPT - LN(SRI) relationship only for the Aircrew Life Support (AFSC 122x0) and AGE Mechanic (AFSC 423x5) samples, and Job Experience served as a moderator only in the Avionic Communications (AFSC 328x0), AGE Mechanic (AFSC 423x5) and Personnel (AFSC 732x0) samples. Thus there was only mixed support for the hypothesis that degree of initial task learning, as measured by amount of previous experience, moderates the SRI - task performance relationship. Support also was mixed for the hypothesized moderating effect of incumbent cognitive ability -- statistically significant moderator effects were found only in the Air Traffic Control Operator (AFSC 272x0), Avionics Communications (AFSC 328x0), and Personnel (AFSC 732x0) samples. On the other hand, much stronger support was found for the hypothesized moderating effect of task difficulty on the SRI - task performance relationship -- statistically significant interaction effects on WTPT scores were supported in five of the eight AFSs.

To explore the form of the interaction effects that were statistically significant, we examined the unstandardized regression results for the full regression models (i.e., those including statistically significant cross-product terms), using procedures described by Jaccard, Turrisi, and Wan (1990). Specifically, we selected high (HI) and low (LO) values for LN(SRI) and each of the moderators as the mean, plus and minus two standard deviations, respectively. We substituted these HI and LO values for LN(SRI) and each moderator into the appropriate full unstandardized regression equations to calculate and plot predicted WTPT values. Figure 1 shows plots that are representative of the forms of the interactions that were found to be statistically significant.

TABLE 3: Summary of Correlational and Hierarchical Moderated Regression Results

AFSC	NOBS	Task r (WTPT,SRI)	Job Experience	Task Experience	Aptitude	Difficulty
122x0	1591	-0.17**	0.00339*	0.00139	0.00049	0.00935**
272x0	1548	-0.32**	0.00000	0.00002	0.00166*	0.01499**
324x0	1990	-0.09**	0.00025	0.00049	0.00072	0.00394**
328x0	1409	-0.06*	0.00002	0.00364*	0.00429**	0.00162
423x5	3495	-0.22**	0.00653**	0.00179**	0.00031	0.01529**
426x2	1123	-0.06**	0.00101	0.00000	0.00001	0.00000
492x1	872	0.12**	0.00176	0.00074	0.00016	0.00037
732x0	1928	-0.13**	0.00068	0.00326**	0.00540**	0.00842**

*$p < 0.05$, ** $p < 0.01$.
Note. AFSC = Air Force Specialty Code; NOBS = Number of Observations; WTPT = Walk Through Performance Test Score; SRI = Skill Retention Interval.

Figure 1a plots the LN(SRI) x Task Experience interaction, and shows that the length of the SRI had a dramatically adverse effect on task performance for individuals with low task experience, whereas individuals with high task experience maintained a high level of task performance regardless of how long it had been since they performed the task. This finding is consistent with the Driskell et al.'s (1992) meta-analytic findings that performance decrements may not readily incur for tasks that are well learned initially. Similar results were found with respect to the LN(SRI) x Job Experience interaction (see Figure 1b). Once again, individuals with higher (general) job experience exhibited relatively high task performance regardless of how long it had been since they performed the tasks, whereas lesser-experienced individuals' task performance suffered dramatically with extended SRIs. In addition to the task learning effects suggested by Figure 1a, these results may also reflect additional opportunities of longer-tenured incumbents to gain expertise with a broader array of tasks as compared to shorter-tenured incumbents (Ford, Quinones, Sego, & Sorra, 1992).

FIGURE1: Representative patterns of statistically significant interactions supported in Table 3.

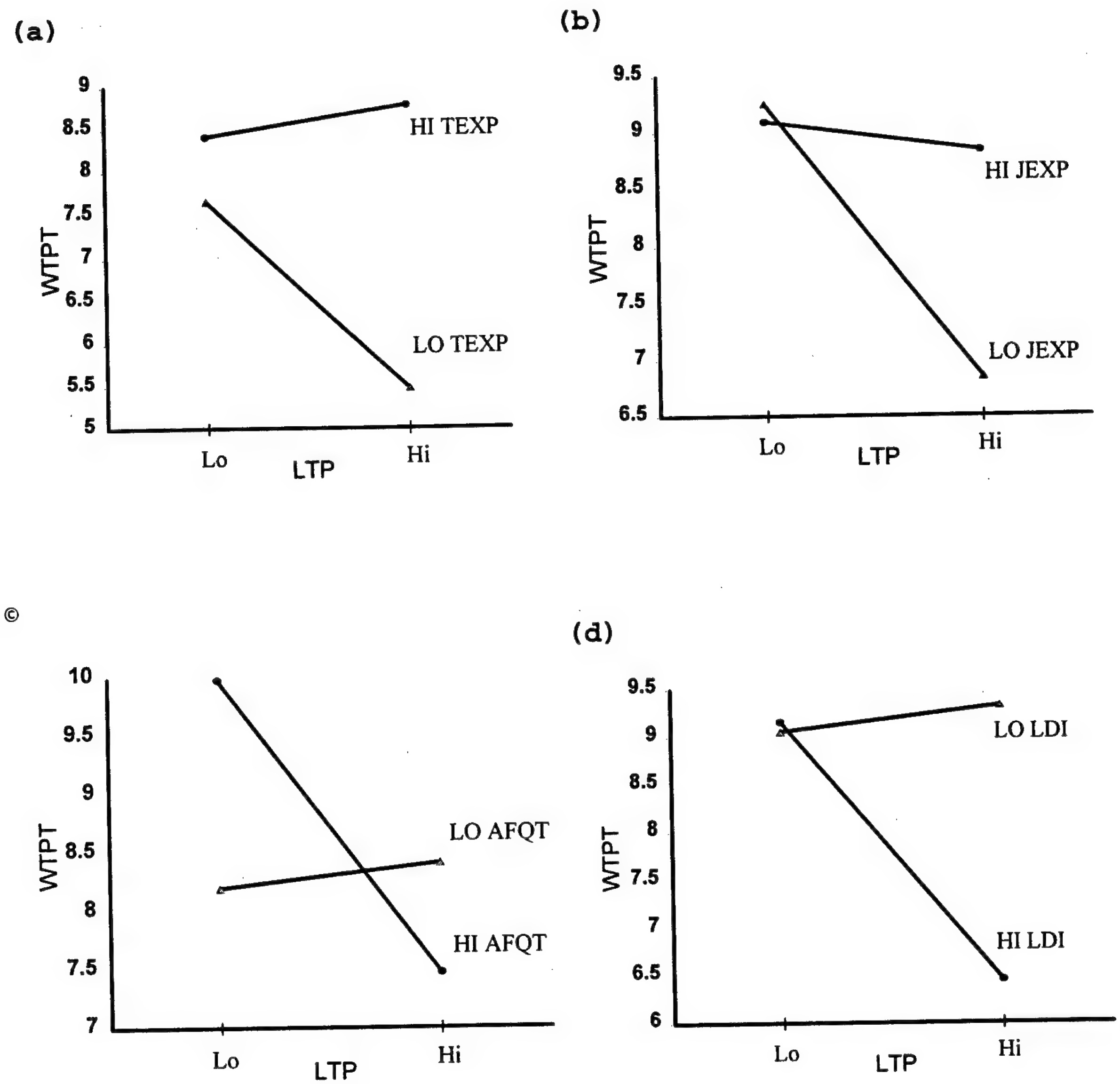


Figure 1c plots the LN(SRI) x Aptitude interaction. Results generally show that higher ability incumbents outperform lower ability incumbents on tasks that both groups had performed recently, but that with extended SRIs both higher- and lower-ability incumbents performed relatively poorly.

Finally, Figure 1d shows the form of LN(SRI) x Task Difficulty interaction. The effects here are clear: performance on less difficult tasks remains high even after extended SRIs, whereas performance decreases dramatically over extended SRIs for more difficult tasks. As predicted, those tasks that are initially the most difficult to learn are also those that are most easily forgotten.

DISCUSSION

The primary goals of this study were to test theoretically based moderator effects on the relationship between extended skill retention intervals and decrements in task performance in realistic work-related settings. Our intent was to corroborate earlier findings obtained in (primarily) laboratory settings and to evaluate the external validity of these findings in an applied field context. Results provided mixed support for hypotheses.

First, the negative relationship between length of SRI and task performance that has been found repeatedly elsewhere (e.g., Arthur et al., 1995; Farr, 1987) also received strong support here: statistically significant negative SRI - task performance relationships were found in 7 of the 8 samples studied. Thus these findings, obtained from realistic work-related settings, corroborate and support the external validity of similar findings obtained previously in (mainly) laboratory studies. However, SRI- task performance relationships were not strong. At most, about 10% of the variance in task performance was accounted for by length of the SRI. So why wasn't this allegedly robust relationship stronger?

There are perhaps four reasons. First, since almost all research participants were first-term airmen, there may have been restricted range on SRI as compared to more extended SRIs that may exist among longer-tenured incumbents. Still, the mean SRIs reported here are generally much longer than those reported in studies reviewed by Arthur et al. (1995) and Farr (1987). Second, restriction of range on ability due to preselection on ASVAB scores could have attenuated the SRI - task performance relationship. Range restriction, and even indirect range restriction as in this case, can attenuate empirical associations, and there are a number of convincing rationales for correcting for range restriction when critical values (e.g., unrestricted predictor and criterion standard deviations) are available or estimable (Linn, Harnish, Dunbar, 1981; Thorndike, 1949). However, we chose not to correct for range restriction due to preselection under the rationale that the present findings would most meaningfully be generalizable to other, similarly preselected (and consequently restricted) populations. Third,

experience performing similar, related tasks on the job may have facilitated recall of procedural task steps even though the examinee may not have performed the particular tasks that were tested in the WTPT for some time. Thus, for example, although some participants may not have installed electronic component "X" in some time, their more recent experience installing other electronic components may have facilitated their installation of component "X" in the testing situation. This problem is often controlled for in laboratory studies by using novel, contrived tasks, but points to a potential confound that is much more difficult to control in realistic work settings. Finally, additional unmeasured performance determinants (e.g., motivation, prior job knowledge) may have overshadowed the effects of extended SRIs on task performance.

Second, moderator hypotheses also received only mixed support. This may partly reflect general difficulties in detecting interactions in nonexperimental data (e.g., McClelland & Judd, 1993). It may also reflect aspects of the measurement of constructs in this study or the influences of variables not measured here. For example, we operationalized the degree of initial skill learning using surrogate task and job experience measures. While we are convinced that these measures reflect aspects of initial skill learning, at least as the SRI commences, we would encourage future research to employ more direct assessments of initial skill learning, when construct-valid measures are available, as well as these surrogate indices. Second, overlearning is related to skill maintenance over periods of skill non-use (Driskell et al., 1992). While our measures of task experience reflect the amount, or activity level for each task (Ford et al., 1992; Quinones, Ford, & Teachout, 1995) it may be the amount of task overlearning rather than the specific activity level for a task that is more important in skill maintenance. Similarly, other aspects of the training program and work climate may have facilitated skill maintenance and training transfer. For example, in-training transfer enhancing activities such as stimulus variability, physical and psychological fidelity, and teaching of principles and meaningfulness may have occurred to facilitate skill maintenance (Baldwin & Ford, 1988; McGehee & Thayer, 1961; Thayer & Teachout, 1995). A favorable work climate may also have facilitated transfer, reinforced learning, and skill maintenance (Baldwin & Ford, 1988; Rouiller & Goldstein, 1993; Thayer & Teachout, 1995; Tracey, Tannenbaum, & Kavanagh, 1995). Nevertheless, and despite these potential confounds of more naturalistic, applied research settings, our results provided support, albeit mixed, for theoretically-based moderator hypotheses.

Despite the fact that our findings were not unequivocally supportive of hypotheses, we see potential practical implications of this stream of research. For example, results showed that performance remained essentially intact on less (as compared to more) difficult tasks even after extended SRIs. Thus, it may be more practical to target overtraining or refresher training more specifically at those job tasks that are more difficult to learn, and on which performance decrements are more likely to incur, than to overtrain or offer refresher training less discriminantly on all job tasks. Of course, some independent assessment of task learning difficulty such as the job-analysis based measures reported here would be useful in identifying tasks as targets for refresher training, but other data such as passing rates or training times allocated for various training modules may also be useful for this purpose. Results also suggest

that training resources could be best targeted as relatively less experienced and cognitively able employees. Training allocation decisions such as these will become even more important as organizations continue to face increasingly tight training budgets and strong demands for accountability. Thus, one of the most critical needs at this juncture is for additional field research on the SRI - performance decrement relationship such as is reported here, to determine the extent to which, and under what conditions, earlier laboratory findings generalize to actual work settings.

REFERENCES

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. Journal of Applied Psychology, 77, 598-614.

Ackerman, P. L., & Humphreys, L. G. (1991). Individual differences theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol 1.) (pp. 223-282). Palo Alto, CA: Consulting Psychologists.

Adams, J. A. (1972). Research and the future of engineering psychology. American Psychologist, 27, 615-622.

Arthur, W. Jr., Bennett, W. Jr., McNelly, T., & Stanush, P. I. (1995). Skill retention and decay: A meta-analysis. (Manuscript submitted for publication).

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63-105.

Burtch, L. D., Lipscomb, M. S., & Wissman, D. J. (1982). Aptitude requirements based on task difficulty: Methodology for evaluation. (AFHRL-TR-81-34). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Carretta, T. R., & Ree, M. J. (1995). Air Force Officer Qualifying test validity for predicting pilot training performance. Journal of Business and Psychology, 9, 379-388.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. NY: Wiley.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis for field settings. Boston: Houghton Mifflin.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi experimentation. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol. 1)(pp. 491-576). Palo Alto, CA: Consulting Psychologists.

Department of Defense. (1984, July). Test manual for the armed services aptitude battery. North Chicago, IL: United States Military Entrance Processing Command.

Dobbins, G. H., Lane, I. M., & Steiner, D. D. (1988). A note on the role of laboratory methodologies in applied behavioural research: Don't throw the baby out with the bath water. Journal of Organizational Behavior, 9, 281-286.

Driskell, J. E., & Salas, E. (1992). Can you study real teams in contrived settings? The value of small group research to understanding teams. In R. W. Sweeny and E. Salas (Eds.), Teams: Their training and performance, (pp. 101-123). Norwood, NJ: Ablex.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. Journal of Applied Psychology, 77, 615-622.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. Educational and Psychological Measurement, 52, 721-725.

Farr, J. (1987). The long-term retention of knowledge and skills. NY: Springer-Verlag.

Fendrich, D. W., Healy, A. F., Meiskey, L., Crutcher, R. J., Little, W., & Bourne, L. E. Jr. (1988). Skill maintenance: Literature review and theoretical analysis. (AFHRL-TP-87-73). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Ford, J. K. Quinones, M. A., Sego, D. J., & Sorra, J. (1992). Factors affecting the opportunity to perform trained tasks on the job. Personnel Psychology, 45, 511-527.

Gagne, R. M., Baker, K. E., & Foster, H. (1950). On the relation between similarity and transfer of training in the learning of discriminative motor tasks. Psychological Review, 57, 67-79.

Gallagher, J. J. (1994). Teaching and learning: New models. Annual Review of Psychology, 45, 171-195.

Goldstein, I. L. (Ed.) (1989). Training and development in organizations. San Francisco, CA: Jossey-Bass.

Goldstein, I. L. (1991). Training in work organizations. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol 2.) (pp. 507-520). Palo Alto, CA: Consulting Psychologists.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-213.

Hedge, J. W., & Teachout, M. S. (1986). Job performance measurement: A systematic program of research and development. (AFHRL-TP-86-37). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., & Teachout, M. S. (1992). An interview approach to work sample criterion measurement. Journal of Applied Psychology, 77, 453-461.

Herrnstein, R. J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. NY: Free Press.

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. Annual Review of Psychology, 41, 109-140.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Interaction effects in multiple regression. Newbury Park, CA: Sage.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal analysis: Models, assumptions, and data. Beverly Hills, CA: Sage.

Kavanagh, M. J., Borman, W. C., Hedge, J. W., & Gould, R. B. (1987). Job performance measurement in the military: A classification scheme, literature review, and directions for research. (AFHRL-TR-87-15). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology, 78, 311-328.

Lance, C. E., Hedge, J. W., & Alley, W. E. (1989). Joint relationships of task proficiency with aptitude, experience, and task difficulty: A cross-level interactional study. Human Performance, 2, 249-272.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. Journal of Applied Psychology, 77, 437-452.

Latham, G. P., & Lee, T. W. (1986). Goal setting. In E. A. Locke (Ed.), Generalizing from laboratory to field settings. Lexington, MA: Lexington.

Laue, F. J., Hedge, J. W., Wall, M. L., Pederson, L. A., & Bentley, B. A. (1992). Job performance measurement system development process. (AL-TR-1992-0120). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Technical Training

Research Division.

Laue, F. J., Teachout, M. S., & Harville, D. L. (1990, August). Assessing the substitutability of surrogates (sic) measures of job performance for hands-on work sample tests. Paper presented at the meeting of the American Psychological Association, Boston, MA.

Linn, R. L., Harnish, D. L., & Dunbar, S. (1981). Corrections for range restriction: An empirical investigation of conditions resulting in conservative corrections. Journal of Applied Psychology, 66, 655-663.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114, 376-390.

McGehee, W., & Thayer, P. W. (1961). Training in business and industry. New York: Wiley.

Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The forgetting of instrument flying skills. Human Factors, 13, 397-405.

Michalak, D. F. (1981). The neglected half of training. Training and Development Journal, 35(5), 22-28.

Mumford, M, D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1987). Measuring occupational difficulty: A construct validation against training criteria. Journal of Applied Psychology, 72, 578-587.

Naylor, J. C., & Briggs, G. E. (1961). Long-term retention of learned skills: A review of the literature. (ASD-TR-61-390). Wright Patterson AFB, OH: Aerospace Medical Laboratory, Aeronautical Systems Division.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. Journal of Applied Psychology, 79, 845-851.

Prophet, W. W. (1977). Long-term retention of flying skills: A review of the literature. (AD/A-036-077). Alexandria, VA: Human Resources Research Organization.

Quinones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. Personnel Psychology, 48, 887-910.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. Educational and Psychological Measurement, 54, 459-463.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. Journal of Applied Psychology, 80, 721-730.

Ree, M. J., & Earles, J. A. (1991a). Predicting training success: Not much more than g. Personnel Psychology, 44, 321-332.

Ree, M. J., & Earles, J. A. (1991b). The stability of g across different methods of estimation. Intelligence, 15, 271-278.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. Journal of Applied Psychology, 79, 518-524.

Rouiller, J. Z., & Goldstein, I. L. (1993). The relationships between organizational transfer climate and positive transfer of training. Human Resource Development Quarterly, 4, 377-390.

Royer, J. M. (1979). Theories of the transfer of learning. Educational Psychologist, 14, 53-69.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. Journal of Applied Psychology, 67, 605-610.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: Review. (TP 313). Alexandria, VA: U.S. Army Research Institute.

Schmidt, F. J., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. Journal of Applied Psychology, 73, 46-57.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. Annual Review of Psychology, 43, 583-626.

Tannenbaum, S., & Yukl, G. (1992). Training and development in work organizations. Annual Review of Psychology, 43, 399-442.

Teachout, M. S., & Pellum, M. W. (1991). Air Force research to link standards for enlistment to on-the-job performance. (AFHRL-TR-90-90). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Thayer, P. W., & Teachout, M. S. (1995). A climate for transfer model. (AL/HR-TP-95-0035). Brooks AFB, TX: Armstrong Laboratory, Technical Training Research Division.

Thorndike, R. (1949). Personnel Selection. New York: Wiley.

Tracey, J. B., Tannenbaum, S. I., & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of work environment. Journal of Applied Psychology, 80, 239-252.

Weeks, J. (1984). Occupational learning difficulty: A standard for determining the order of aptitude requirement minimums. (AFHRL-SR-84-26). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Weiss, H. M. (1991). Learning theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol 1.) (pp. 171-222). Palo Alto, CA: Consulting Psychologists.

Wigdor, A. K., & Green, B. F. Jr. (Eds.). (1986). Assessing the performance of enlisted personnel: Evaluation of a joint-service research project. Washington, DC: National Academy Press.

Wigdor, A. K., & Green, B. F. Jr. (Eds.). (1991). Performance assessment for the workplace. Washington, DC: National Academy Press.

Wisher, R. A., Sabol, M. A., Hillel, H., & Kern, R. P. (1991). Individual ready reserve (IRR) call-up: Skill decay. ARI Research Report No. 1595. Alexandria, VA: U.S. Army Research Institute.

FOOTNOTES

[1]Matrices of intercorrelations among all study variables are available from the first author.

[2]As we explain later, we chose not to correct data for range restriction, as it was our intent to make inferences to similarly restricted populations rather than to hypothetical unrestricted populations.